Quantifying Uncertainty through Global and Mesoscale Ensembles

Teddy R. Holt
Naval Research Laboratory
7 Grace Hopper Ave
Monterey CA 93943-5502
phone: (831) 656-4740 fax: (831) 656-4769 e-mail: holt@nrlmry.navy.mil

Document Number: N0001409WX20032
http://www.nrlmry.navy.mil/

LONG TERM GOALS

The long-term goal of this project is to develop robust global and mesoscale ensemble analysis and forecast systems that are able to provide probabilistic forecast guidance in an operational environment. Although current guidance relies primarily on single deterministic forecasts from numerical weather prediction models, recent advances in probabilistic prediction, or ensemble forecasting, have made this technique a very powerful tool for providing more complete input on environmental conditions, including a measure of the forecast confidence. Probabilistic prediction of Navy relevant high-impact weather will significantly benefit sea strike and sea shield functions if there are clearly identified *user-relevant* norms. Ensembles also allow for the covariance between relevant weather variables to be taken into account, for example, indicating that *if* the wind speed is high, *then* visibility will also be high.

OBJECTIVES

The objective of this project is to incrementally develop and test an integrated multi-scale ensemble modeling framework and forecasting system, and validate and transition it to operations at the Fleet Numerical Meteorology and Oceanography Center (FNMOC). This system will be a state-of-the-art global and mesoscale ensemble forecasting system using the Navy Operational Global Atmospheric Prediction System (NOGAPS) and the Coupled/Ocean Atmosphere Mesoscale Prediction System (COAMPS®[1]). It will incorporate consideration of uncertainty in the initial state (and lateral boundaries where appropriate) along with consideration of uncertainty in the model formulation. It will be a user-friendly, easily re-locatable system flexible enough to incorporate new advances in data assimilation and post-processing schemes. The system will provide high fidelity, dynamically consistent probabilistic forecasts, and estimates of forecast uncertainty for the battlespace environment that will directly improve tactical decision aids (TDA) and applications related to homeland security and the global war on terrorism.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **30 SEP 2009**	2. REPORT TYPE **Annual**	3. DATES COVERED **00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE **Quantifying Uncertainty Through Global And Mesoscale Ensembles**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,7 Grace Hopper Ave,Monterey,CA,93943**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Code 1 only

14. ABSTRACT
The long-term goal of this project is to develop robust global and mesoscale ensemble analysis and forecast systems that are able to provide probabilistic forecast guidance in an operational environment. Although current guidance relies primarily on single deterministic forecasts from numerical weather prediction models, recent advances in probabilistic prediction, or ensemble forecasting, have made this technique a very powerful tool for providing more complete input on environmental conditions, including a measure of the forecast confidence. Probabilistic prediction of Navy relevant high-impact weather will significantly benefit sea strike and sea shield functions if there are clearly identified user-relevant norms. Ensembles also allow for the covariance between relevant weather variables to be taken into account, for example, indicating that if the wind speed is high, then visibility will also be high.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**9**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

APPROACH

A global Ensemble Transform (ET) ensemble system has been developed as part of research performed in recent NRL base projects on global modeling and predictability (Bishop and Toth 1999, McLay et al. 2007). In addition, an ET mesoscale ensemble system, driven by this global ensemble, has been developed within the FY06-09 NRL 6.2 project "Probabilistic Prediction of High-Impact Weather" (Bishop et al. 2009). The outcomes of those projects have been incorporated into this Rapid Transition Process (RTP) project, allowing us to use the combined ONR and PMW-120 funding to move rapidly forward with both the global and mesoscale ensemble research systems to fully develop, test, validate, and operationalize a multi-scale ensemble forecasting capability at FNMOC. The key individuals working on this project are Carolyn Reynolds and Justin McLay (global ensembles), and Teddy Holt and Craig Bishop (mesoscale ensembles).

This next-generation, multi-scale ensemble system will consist of (a) the representation of initial and model uncertainty in NOGAPS, (b) the representation of initial uncertainty, lateral boundary condition (LBC) uncertainty, and model uncertainty in COAMPS, (c) an integrated, robust global-mesoscale ensemble infrastructure (to ensure consistent perturbations between the global and mesoscale systems, result in improvements to the global system having an immediate impact on the mesoscale system, and promote ease of use), (d) an ensemble verification system, including a scorecard for objectively evaluating overall skill, (e) a demonstration of performance capabilities in an operational setting, and (f) an ensemble post-processing scheme designed to ensure reliable probabilistic forecasts to remove biases from the forecasts of individual members together with biases in the ensemble spread. The system will be tested using scientific studies and norms and comprehensive ensemble forecast experiments.

WORK COMPLETED

For the development of the global ensemble system, a comparison between the ET (Bishop and Toth 1999) and the previously operational Bred Vector (BV, Toth and Kalnay 1993) scheme has shown the ET scheme to be superior under many different metrics (McLay et al. 2008). The ET scheme was transitioned to FNMOC in March 2008. Research is now underway to improve the current fit of the initial perturbations to the analysis error variance estimates produced by the data assimilation scheme using a “banded” ET which should be transitioned to operations shortly. Research also investigated the impact of including model uncertainty through inclusion of parameter variations as well as incorporation of a simple model of diurnal SST variations (Zheng and Beljaars 2005).

For the development of the mesoscale ensemble system, the ET method for generating initial perturbations implemented within COAMPS in 2008 has been extensively tested in FY09 using the Joint Mesoscale Experiment (JME) domain (45-, 15-km; 33 members) over Korea, with a focus on developing and improving post-processing techniques. Four bias-correction techniques (static, Kalman-Filter (KF), running-mean, and weighted running mean) and calibration (ensemble spread) techniques have been developed and tested as well as new filtered model output statistics (FMOS) and a new approach to Bayesian Model Averaging (BMA) ensemble post-processing.

RESULTS

a. Global ensemble system: While the first version of the global ET has already been transitioned to operations, there are still issues with the current ET implementation that merit further research. One issue is that the ET scheme has too little ensemble variance in the tropics when compared to the analysis error variance estimate that is produced by the operational data assimilation system. Recently a new "banded" ET in which the transform is performed in latitudinal bands has been developed to mitigate these problems. The new banded ET results in a much closer match to the analysis error variance estimates and also gives improved performance under a variety of metrics. Figure 1 shows two examples of this improved performance, including reduced RMS error of the NH Midlatitude 500-hPa ensemble mean RMS error, and less redundancy between ensemble members at 120-h indicated by higher values of the normalized eigenvalues of the ensemble covariance matrix.

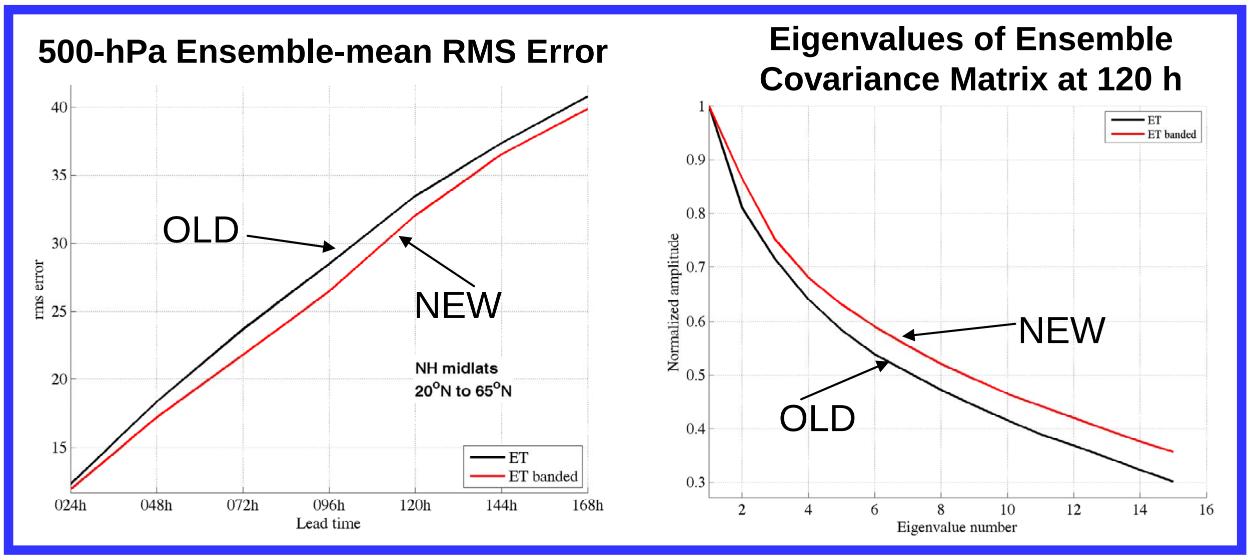

Fig. 1. Results for the operational version of the ET (black curves) as the new banded ET (red curves). The left panel shows ensemble 500-hPa geopotential height ensemble-mean forecast error for 20-65N, as a function of forecast time. The right panel shows the normalized eigenvalues of the ensemble covariance matrix at 120 h.

Another way to improve the current ET implementation is through inclusion of model uncertainty. Experiments combining the ET with parameter variations in just the convective scheme are compared with parameter variations in both the convective and boundary layer scheme. It is found that while parameter variations in the convective scheme improve performance in the tropics, they have little impact on mid-latitudes. In contrast, variations in both the convective and boundary layer parameters provide improved performance in both the tropics and the mid-latitudes. Figure 2 shows improved performance in the ability of the ensemble variance to differentiate between a wider range of ensemble mean forecast errors when both the convective and boundary layer parameters are varied (green curve) as opposed to just varying the convective parameters (cyan curve) or the control (dark blue curve). Experiments are also being carried out with the addition of a simple model of diurnal SST variations following Zheng and Beljaars (2005). Experiments show mostly neutral to improved performance in

standard ensemble metrics (e.g., improved relationship between ensemble variance and forecast error variance, cyan curve in Figure 2), as well as improved simulation of equatorial waves such as Kelvin waves (not shown), which should improve prediction of the Madden Julian Oscillation.

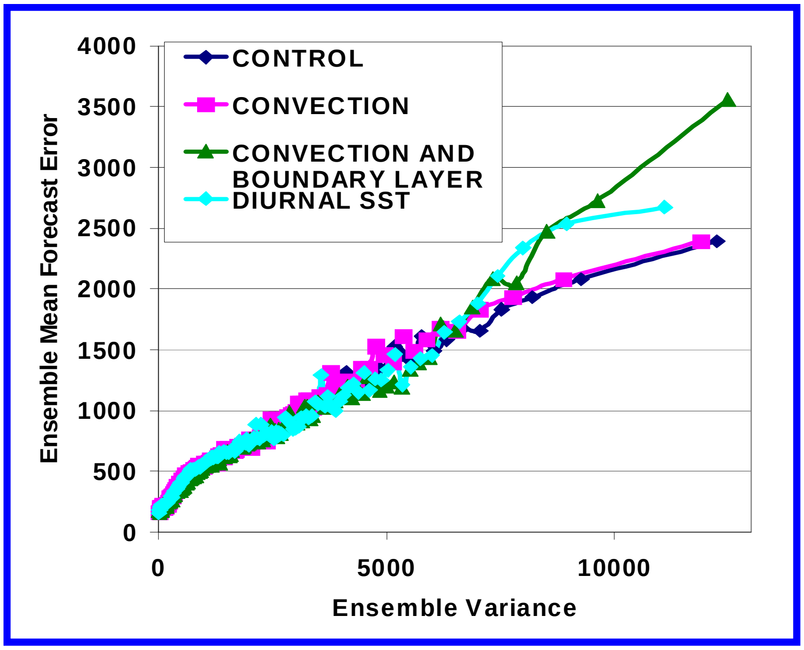

Fig. 2. Ensemble mean forecast error variance compared to ensemble variance for NH midlatitude 500-hPa height 48-h forecasts, May-June 2007. Dark blue: Control ensemble (no model uncertainty). Magenta: Variations in convective parameters. Cyan: Diurnal SST variations. Green: Variations in convective and boundary layer parameters.

b. *Mesoscale ensemble system:* New post-processing techniques have been developed and tested using the JME domain for the 3-month time period (12 February to 15 May 2009) using 340 fixed surface observation stations for air-temperature, wind speed, and dew point depression. Figure 3 shows an example of the improvement that KF bias correction affords over no bias correction (raw) for 2-m air temperature. Generally, weighted running mean and KF bias correction perform better than either running mean or static bias correction.

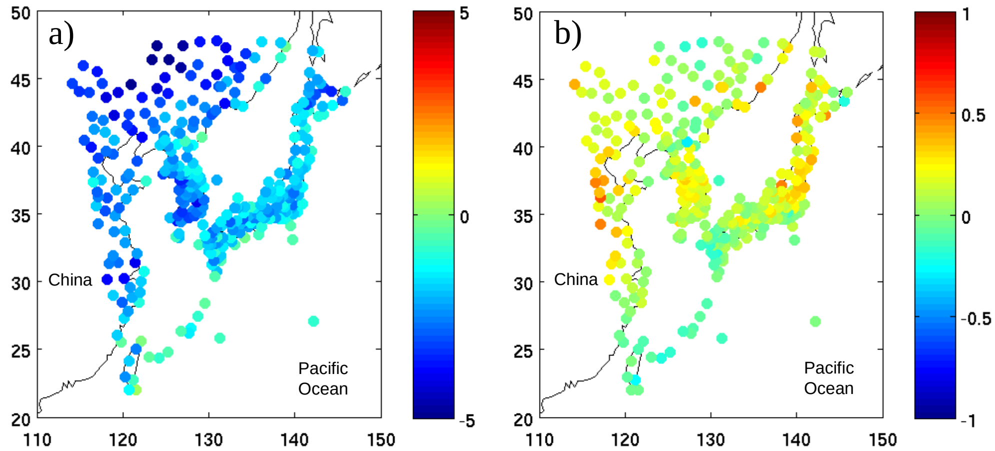

Fig. 3. COAMPS mean 48-h forecast bias of 2-m air temperature (K) for the observations in the 15-km JME domain for 15 April to 15 May 2009 for a) raw (no correction), and b) Kalman-Filter correction.

In addition to bias correction, calibration of second moments (spread) have also been developed, with the second moment estimated as a linear function of the ensemble sample covariance. Figure 4 shows the calibrated Taligrand diagram and reliability for the KF-calibrated 48-h forecasts of 2-m air temperature. The probability integral diagram (Fig. 4a) shows an approximately flat distribution as desired, in contrast to an under-dispersive non-calibrated distribution (Figure not shown). Likewise the reliability diagram (Fig. 4b) for the median air temperature (blue line), 10 percentile (red), 30 percentile (green), 70 percentile (cyan) and 90 percentile (magenta) all show excellent agreement of forecast probability with relative frequency across all ranges.

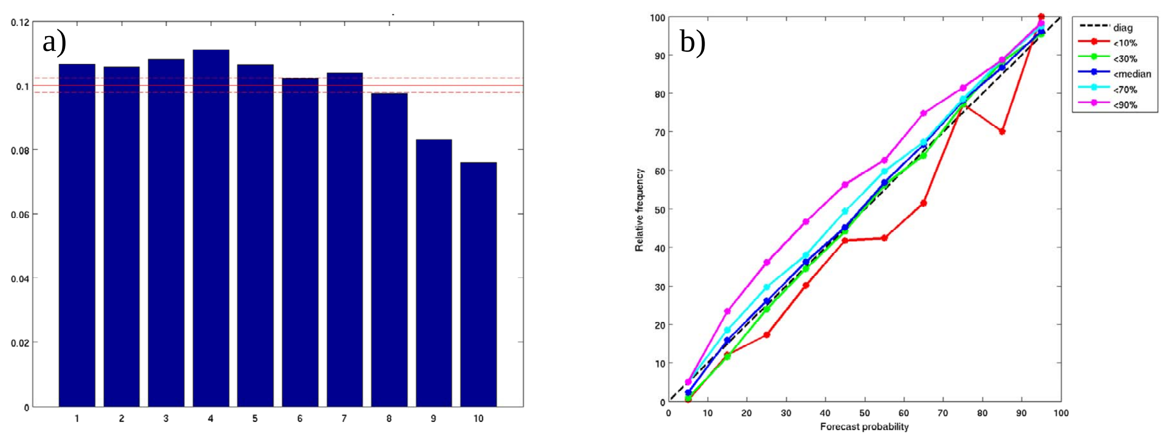

Fig. 4. COAMPS 15-km JME 48-h KF-calibrated 2-m air temperature (K) forecasts from 12 February to 9 June 2009 for a) Probability integral transform (PIT) histogram, and b) reliability.

The newly developed weather forecasting tool (FMOS) improves existing statistical corrections of systematic forecast error, particularly when the training data set is small (Bishop 2009, US provisional patent 68/181686). Previous MOS systems (i.e., Wilson and Vallée 2002) use blended regression

equations based on earlier versions of the model with equations based on more recent versions of the model, assuming that significant model changes are typically separated by more than two years. This assumption is not satisfied for US Navy operations in which model domains are typically deployed in a specific theatre for a period between two weeks and two years. In principal, new MOS equations need to be developed for each unique theatre in which the forecasting model is deployed. Hence, the Navy's deployment periods are too short for the estimated regression coefficients to stabilize. Any attempt to estimate the regression coefficients using MOS with such a small data set will result in very noisy regression coefficients. The new FMOS developed here stabilizes and spatially smoothes such noisy regression coefficients associated with small training data sets by optimally combining a data-poor MOS estimate with a smooth *prior* estimate. MOS methods assume that on the kth day/event, the ith model forecast variable f_i^k is related to the corresponding verifying analysis/observation variable y_i^k via the stochastic process $f_i^k = a_i y_i^k + b_i + \varepsilon_i^k$ where ε_i^k is a random number with mean zero and variance σ_r^2 that is statistically independent of y_i^k. Fig. 5a gives an example of the estimates a_i^{MOS} of a_i one would obtain using the standard MOS equations and simulated meteorological data when just 16 consecutive meteorological events were used to make the estimate. Fig. 5c gives the corresponding true values. The beneficial effect of the FMOS tool is evident in the comparison of Fig. 5a and 5e indicating that while the MOS estimate of the regression coefficient is noisy, the FMOS estimate is smooth. Comparison of the contour scales for Fig. 5b and 5f shows that the maximum errors incurred by the MOS estimate are more than 5 times larger than those incurred by the FMOS estimate, and the mean square error (mse) of the MOS estimate is 21 times larger than the mse of the FMOS estimate.

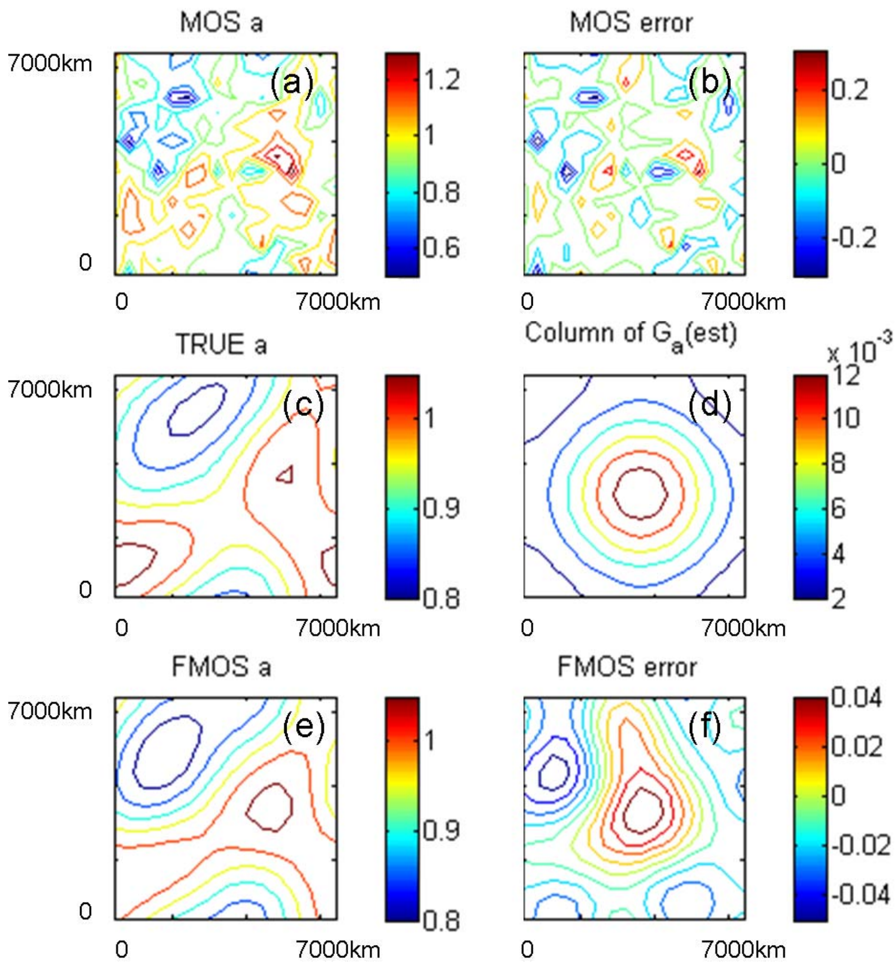

Fig. 5: Comparison of estimates of the slope coefficient a ***from MOS and FMOS from 16 realizations. In (a) and (b) are shown the MOS estimate of a and its error, respectively. The true a-field is shown in (c) while (d) indicates the spatial covariance function of the a-field obtained from a spectral analysis of the MOS estimate shown in (a). The FMOS estimate of a and its error are shown in (e) and (f), respectively.***

The advantages of FMOS over MOS are that (a) its estimates of regression coefficients are more accurate – particularly when the training data set is small, and (b) its estimates of regression coefficients lie between those of a perfect forecast model and those of the actual model. It can be shown that, if properly configured, FMOS will deliver equivalent results to MOS in the limit of an infinite data training set.

Finally, BMA methods of ensemble post-processing in which continuous probability density functions (pdfs) are constructed from ensemble forecasts by centering functions around each of the ensemble members have been examined. The primary purpose of BMA methods is to correct systematic mismatches between ensemble variance and the error variance of the ensemble mean. In Bishop and Shanley (2008), idealized ensemble forecasting experiments were used to show that these methods are liable to produce systematically unreliable probability forecasts of climatologically extreme weather. They showed that the failure of these methods is linked to an assumption that the distribution of truth

given the forecast can be sampled by adding stochastic perturbations to state estimates, even when these state estimates have a realistic climate. The research demonstrates that this assumption is incorrect and argues that such dressing techniques better describe the *likelihood* distribution of historical ensemble mean forecasts given the truth for certain values of the truth. This paradigm shift led to a new approach that incorporates *prior* climatological information into BMA ensemble post-processing via Bayes' theorem. This new approach is shown to cure BMA's ill treatment of extreme weather by providing a *posterior* BMA distribution whose probabilistic forecasts are reliable for both extreme and non-extreme weather forecasts. Work is underway to incorporate this new ensemble post-processing technique into the coupled model ensemble.

IMPACT/APPLICATIONS

Since results indicate improved performance of the global ET system over the previous operational method under a variety of metrics, this new scheme is resulting in more skillful global ensemble forecasts for the Navy. This approach should have a direct impact on current and planned ensemble-based products, such as the probability of gale-force winds, as well as other products that are currently forced by the global ensemble, or will be in the future, including the mesoscale atmospheric ensembles, and surface ocean wave ensembles. In the future, it is hoped that these more skillful probabilistic tools will lead to more applications of probabilistic forecasts for the Department of Defense, including ship and air routing applications.

TRANSITIONS

Based on the superior performance of the ET as compared to the previously-operational bred-vector scheme, the ET scheme was transitioned to the operational global ensemble system in March 2008. It is anticipated that the banded ET will be transitioned to operations in the next few months, and inclusion of model uncertainty transitioned to operations in FY10. In addition, as part of the Navy/Air Force Joint Mesoscale Ensemble (JME) forecasting project we have aided FNMOC in the implementation and execution of a "beta-ops" real-time mesoscale ensemble forecasting system over the Korean peninsula. The COAMPS ET ensemble has been running continuously on the High Performance Computing network's DC3 system since late June 2008. The transition to operations is in progress, with porting to the FNMOC large operational LINUX cluster (OPAL A2) complete. The mesoscale system has been successfully tested by NRL in a quasi-operational configuration at FNMOC using the operational data streams.

RELATED PROJECTS

The global ET and Stochastic Physics research began in preceding years under the NRL base-funded project 6.1-Quantifying Limits of Atmospheric Predictability and the NOAA funded project 6.2 Stochastic Physics. In addition, related research has been done under the NRL base-funded project 6.2-Model Discovery, which uses ensembles to understand model behavior within parameter space, and the NOAA-funded Hurricane Forecast Improvement Project (HFIP). For the mesoscale ensemble development related projects include ONR funded (through the NRL Base program) 6.2-Probabilistic Forecasting of High Impact Weather, 6.2-Hidden Volatility in Environmental State Estimation, and 6.1-Quantifying Limits of Atmospheric Predictability. The research is also related to the NOAA funded projects 6.2 Stochastic Physics and 6.2 Huge Ensemble State Estimation.

REFERENCES

Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748-1765.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Am. Meteor. Soc.*, **74**, 2317-2330.

Wilson, L.J., and M. Vallée, 2002: The Canadian Updateable Model Output Statistics (UMOS) System: Design and Development Tests. *Wea. Forecasting*, **17**, 206–222.

Zheng, X. and A. Beljaars, 2005: A prognostic scheme of sea surface skin temperature for modeling and data assimilation. Geophys. Res. Lett., 32, L14605, doi: 10.1029/2005GL023030.

PUBLICATIONS

Bishop, C. H., T. Holt, J. Nachamkin, S. Chen, J. McLay, J. Doyle, and W. Thompson, 2009: Regional Ensemble Forecasts Using the Ensemble Transform Technique. *Mon. Wea. Rev.*, **137**, 288–298. [published, refereed]

Bishop, C. H. and K. T. Shanley, 2008: Bayesian Model Averaging's problematic treatment of extreme weather and a paradigm shift that fixes it. *Mon. Wea. Rev.* **136**, 4641–4652. [published, refereed]

McLay, J., C. H. Bishop, and C. A. Reynolds, 2009: A local formulation of the ensemble-transform (ET) analysis perturbation scheme. *Wea. Forecasting* [submitted, refereed].

McLay, J., and C. A. Reynolds, 2009: Two alternative implementations of the ensemble-transform (ET) analysis-perturbation scheme: The ET with extended cycling intervals, and the ET without cycling. *Q. J. R. Meteorol. Soc.*, **135**, 1200-1213. [published, refereed]

McLay, J., C. H. Bishop, and C. A. Reynolds, March 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.*, **136**, 1093-1108. [published, refereed]

McLay, J., C. H. Bishop, and C. A. Reynolds, 2007: The ensemble transform scheme adapted for the generation of stochastic perturbations. *Q. J. R. Meteorol. Soc.*, **133**, 1257-1266. [published, refereed]

Reynolds, C. A., J. Teixeira, and J. G. McLay, 2008: Impact of stochastic convection on the ensemble transform. *Mon. Wea. Rev.*, **136**, 4517-4526. [published, refereed]

PATENTS

Bishop, C. H., 2009: Filtered Model Output Statistics, (Approved by NRL Invention Evaluation Board, May, 2009). US provisional patent 68/181686.